

BAB II

TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Enggal Febriani (2017), melakukan penelitian berjudul “Klasifikasi Konten Berita Surat Kabar Berdasarkan Judul Dengan Text Mining Menggunakan Metode Naïve Bayes (Studi Kasus: Radar Kediri)”. Tujuan dari penelitian ini yaitu untuk membangun sistem yang dapat mengkategorikan berita pada surat kabar dengan mengimplementasikan algoritma naïve bayes. Penelitian ini dilatarbelakangi oleh pengamatan peneliti yang memperlihatkan bahwa dalam penyusunan konten berita pada suatu perusahaan dilakukan secara manual sehingga waktu yang digunakan kurang efisien karena sura kabar harus diterbitkan setiap harinya. Kinerja tim redaksi menjadi kurang efektif karena membutuhkan waktu lebih untuk memilah berita secara manual sebelum berita tersebut dicetak pada susunan surat kabar. Hasil yang diperloeh dari penelitian ini yaitu algoritma naive bayes dapat digunakan dalam pengklasifikasian konten berita sehingga membantu tim editor untuk meminimalisir waktu dan mempermudah pekerjaan (Febriani 2017).

Herny Februariyanti, dkk (2012) dalam jurnal Teknologi Informasi DINAMIK melakukan penelitian berjudul “Klasifikasi Dokumen BeritaTeks Bahasa Indonesia menggunakan Ontologi”. Peneitian ini dilatarbelakangi oleh kebutuhan kita akan dokumen seperti dokumen pembelajaran yang begitu banyak bahkan dalam jumlah yang besar terkadang tidak terorganisir maka diperlukan suatu cara untuk mengelompokkan dokumen berdasarkan

topik untuk membuat tumpukan dokumen menjadi terkelompokkan dalam topik-topik yang sesuai. Tahapan yang dilakukan yaitu text processing, text transformation dan pattern discovery (Februariyanti 2012).

Siti Asiyah, dkk (2016) melakukan penelitian berjudul “Klasifikasi Berita Online Menggunakan Metode Support Vector Machine dan K-Nearest Neighbor”. Tujuan dari penelitian ini yaitu untuk mengklasifikasikan berita ke dalam topik yang sesuai. Kemudian melakukan analisa perbandingan performa dari metode yang digunakan. Hasil dari penelitian ini yaitu perbandingan antara 2 metode Support vector Machine dan K-Nearest Neighbor didapatkan hasil bahwa SVM kernel linier lebih baik daripada K-NN (Nur and Fithriasari 2016).

Kesimpulan dari beberapa penelitian terkait yang dijadikan referensi yaitu data mining dapat digunakan untuk pengklasifikasian teks dengan menerapkan berbagai algoritma. Berdasarkan penelitian terdahulu yang dijadikan referensi, perbedaan yang akan dilakukan oleh penulis dalam penelitian ini yaitu penggunaan K-Nearest Neighbor (KNN) dalam pengklasifikasian artikel berita pada portal berita yang akan dibangun menggunakan bahasa pemrograman web python.

2.2. Klasifikasi Text

Klasifikasi teks merupakan proses pengkategorian teks berdasarkan pola atau pattern tertentu yang muncul dan disaring atau ditemukan oleh sistem pengklasifikasian bergantung pada metode apa yang digunakan (Ridok and Latifah 2015). Klasifikasi dokumen bertujuan untuk mengelompokkan dokumen-dokumen yang tidak terstruktur menjadi bersatu dengan kelompok

yang memiliki topik sama serta menggambarkan isi dari dokumen tersebut (Widaningsih and Suheri 2018).

2.3. Content writer

Content writer atau penulis konten merupakan pekerjaan yang saat ini sering dilakukan di dunia maya baik dalam bentuk kontrak atau *freelance*. Penulis konten biasanya akan mengisi rubrik tertentu pada situs web berita atau blog pribadi (Kurniawan, Effendi, and Sitompul 2012). Seorang *content writer* merupakan pekerjaan berjenis tenaga kontrak ataupun *freelance* dari instansi atau perorangan di mana mereka menulis konten berisi karangan, pendapat mengenai suatu barang atau jasa dan laporan seperti berita pada situs web atau blog. Para penulis dunia maya biasanya menulis atau melakukan review dari hal-hal yang viral pada saat itu mulai dari *trending topic*, tren *fashion*, *makeup* dan juga berbagai berita yang bersifat nasional mulai dari kategori bisnis, olahraga, politik dan budaya (Pandrianto and Sukendro 2018).

2.4. K-Nearest Neighbor

K-Nearest Neighbor (kNN) termasuk kelompok instance-based learning. Algoritma ini juga merupakan salah satu teknik lazy learning. kNN dilakukan dengan mencari kelompok k objek dalam data training yang paling dekat (mirip) dengan objek pada data baru atau data testing. K-Nearest Neighbor merupakan salah satu metode yang digunakan dalam pengklasifikasian. Prinsip kerja K-Nearest Neighbor (KNN) adalah mencari jarak terdekat antara data yang akan dievaluasi dengan K tetangga (neighbor) terdekatnya dalam data pelatihan. Teknik

ini termasuk dalam kelompok klasifikasi nonparametric. Di sini kita tidak memperhatikan distribusi dari data yang ingin kita kelompokkan. Teknik ini

